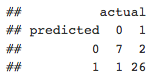
**STACC Task 2 Analysis Report**

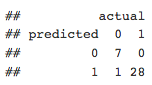
**1. First Confusion Matrix with threshold 0.5 for label = 1**



Observe from the confusion matrix that:

* Out of the 28 label = 1 we classified 26 of them correctly
* Out of the 8 label = 0 we classified 7 of them correctly
* Out of the 36 decisions for SP500 stocks to be in market (1) or not (0) in our test dataset, we **classified** **33 of them** **correctly**

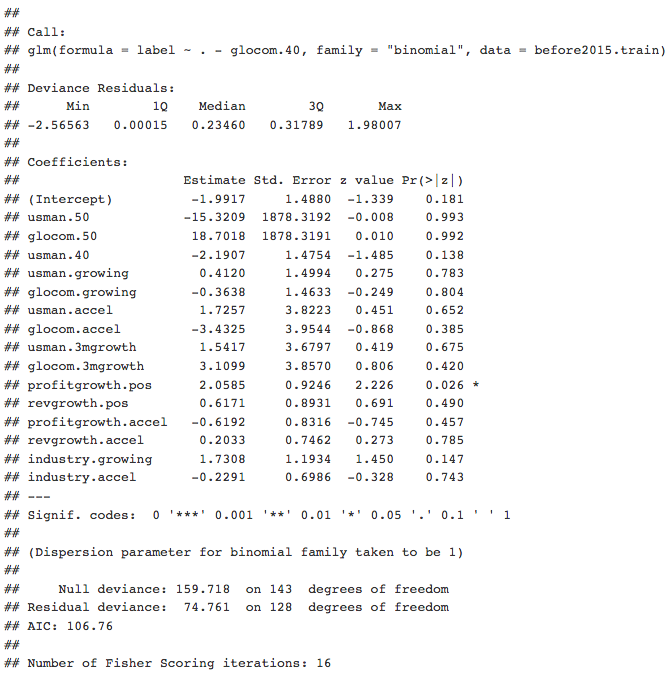
**2. Second Confusion Matrix with threshold 0.3 for label = 1**



We’ve increased accuracy the *number of correct positive cases* from 26 to 28

* Out of the 36 *decisions for SP500 stocks to be in market (1) or not (0)* in our test dataset, we **classified** **35 of them** **correctly** and only on misclassified

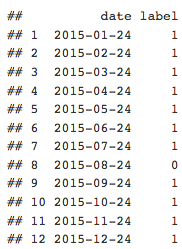
**3. Most Important Variables**

****

Based on summary of our logistic regression model, we found out that variable **usman.50** and **glocom.50** have the highest coefficient value. Following of that, we assume that these two variables are the most important one in our machine learning model.

* usman.50
* glocom.50

**4. Prediction Label for Year 2015**



* Based on the prediction label result above, we should have been in the market in year 2015, except for month of August.